CHAPTER 2

Preliminaries

After a brief review of measure theory, this chapter presents various results about functions of bounded variation, which will play an important role throughout this text.

2.1 Review of Measure Theory

We recall some fundamental ideas from measure theory. Let X be a set. In most cases we will assume that X is a compact metric space.

Definition 2.1.1. A family $\mathfrak{B}$ of subsets of X is called a *σ-algebra* if and only if:
1) $X \in \mathfrak{B}$;
2) for any $B \in \mathfrak{B}$, $\quad X \backslash B \in \mathfrak{B}$;
3) if $B_n \in \mathfrak{B}$, for $n = 1, 2, ...$, then $\bigcup_{n=1}^{\infty} B_n \in \mathfrak{B}$.

Elements of $\mathfrak{B}$ are usually referred to as *measurable sets*.

Definition 2.1.2. A function $\mu : \mathfrak{B} \to \mathbb{R}^+$ is called a *measure* on $\mathfrak{B}$ if and only if:
1) $\mu(\emptyset) = 0$;
2) for any sequence $\{B_n\}$ of disjoint measurable sets, $B_n \in \mathfrak{B}$, $n = 1, 2, ...,$

$$\mu(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n).$$

The triplet $(X, \mathfrak{B}, \mu)$ is called a *measure space*. If $\mu(X) = 1$, we say it is a *normalized measure space* or *probability space*. If X is a countable union of sets of finite μ-measure, then we say that μ is a σ-finite measure. Later on we shall work with probability spaces.

Definition 2.1.3. A family $\mathfrak{A}$ of subsets of X is called an *algebra* if:
1) $X \in \mathfrak{A}$;
2) for any $A \in \mathfrak{A}$, $X \backslash A \in \mathfrak{A}$;
3) for any $A_1, A_2 \in \mathfrak{A}$, $A_1 \cup A_2 \in \mathfrak{A}$.

For any family $\mathfrak{J}$ of subsets of X there exists a smallest σ-algebra, $\mathfrak{B}$, containing $\mathfrak{J}$. We say that $\mathfrak{J}$ generates $\mathfrak{B}$ and write $\mathfrak{B} = \sigma(\mathfrak{J})$.

In practice, when defining a measure μ on a space $(X, \mathfrak{B})$, μ is known only on an algebra $\mathfrak{A}$ generating $\mathfrak{B}$. We would like to know if μ can be extended to a measure on $\mathfrak{B}$. The answer is contained in

Theorem 2.1.1. *Given a set X and an algebra $\mathfrak{A}$ of subsets of X, let μ_1: $\mathfrak{A} \to \mathbb{R}^+$ be a function satisfying $\mu_1(X) = 1$ and*

$$\mu_1(\bigcup_n A_n) = \sum_n \mu_1(A_n)$$

whenever $A_n \in \mathfrak{A}$, for $n = 1, 2, ..., \cup_{n=1}^{\infty} A_n \in \mathfrak{A}$ and $\{A_n\}$ disjoint. Then there exists a unique normalized measure μ on $\mathfrak{B} = \sigma(\mathfrak{A})$ such that $\mu(A) = \mu_1(A)$ whenever $A \in \mathfrak{A}$.

Proposition 2.1.1. *Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space. If $\mathfrak{A}$ is an algebra that generates the σ-algebra $\mathfrak{B}$, then for any $B \in \mathfrak{B}$ and $\varepsilon > 0$ there exists $A \in \mathfrak{A}$ such that $\mu(A \Delta B) < \varepsilon$, where $A \Delta B = (A \backslash B) \cup (B \backslash A)$ is the symmetric difference of A and B.*

Definition 2.1.4. A family $\mathcal{P}$ of subsets of X is called a *π-system* if and only if for any A, B in $\mathcal{P}$ their intersection $A \cap B$ is also in $\mathcal{P}$.

We shall often refer to the following uniqueness theorem [Billingsley, 1968]:

Theorem 2.1.2. *Let $\mathcal{P}$ be a π-system and $\mathfrak{B} = \sigma(\mathcal{P})$. If μ_1 and μ_2 are measures on $\mathfrak{B}$ and $\mu_1(B) = \mu_2(B)$ for any $B \in \mathcal{P}$, then $\mu_1 = \mu_2$.*

Definition 2.1.5. Let X be a topological space. Let $\mathfrak{O}$ denote a family of open sets of X. Then the σ-algebra $\mathfrak{B} = \sigma(\mathfrak{O})$ is called the *Borel σ-algebra* of X and its elements, *Borel subsets* of X.

Definition 2.1.6. Let $(X, \mathfrak{B}, \mu)$ be a measure space. The function $f : X \to \mathbb{R}$ is said to be *measurable* if for all $c \in \mathbb{R}$, $f^{-1}(c, \infty) \in \mathfrak{B}$, or, equivalently, if $f^{-1}(A) \in \mathfrak{B}$ for any Borel set $A \subset \mathbb{R}$.

If X is a topological space and $\mathfrak{B}$ is the σ-algebra of Borel subsets X, then each continuous function $f : X \to \mathbb{R}$ is measurable.

Definition 2.1.7. Let $\mathfrak{B}_n$ be a σ-algebra, $n = 1, 2 \ldots.$ Let $n_1 < n_2 < ... < n_r$ be integers and $A_{n_i} \in \mathfrak{B}_{n_i}$, $i = 1, ..., r$. We define a *cylinder set* to be a set of the form

$$C(A_{n_1}, ..., A_{n_r}) = \{\{x_j\} \in X : x_{n_i} \in A_{n_i}, \quad 1 \leq i \leq r\}.$$

Definition 2.1.8. (Direct Product of Measure Spaces)
Let $(X_i, \mathfrak{B}_i, \mu_i)$, $i \in \mathbb{Z}$ be normalized measure spaces. The direct product measure space $(X, \mathfrak{B}, \mu) = \Pi_{i=-\infty}^{\infty}(X_i, B_i, \mu_i)$ is defined by

$$X = \Pi_{i=-\infty}^{\infty} X_i \quad \text{and} \quad \mu(C(A_{n_1}, ..., A_{n_r})) = \Pi_{i=1}^{r} \mu_{n_i}(A_{n_i}).$$

It is easy to see that finite unions of cylinders form an algebra of subsets of X. By Theorem 2.1.1 it can be uniquely extended to a measure on $\mathfrak{B}$, the smallest σ-algebra containing all cylinders.

2.2 Spaces of Functions and Measures

Let $\mathfrak{F}$ be a linear space. A function $\|\cdot\| : \mathfrak{F} \to \mathbb{R}^+$ is called a *norm* if it has the following properties:

$$\|f\| = 0 \Leftrightarrow f \equiv 0$$

$$\|\alpha f\| = |\ \alpha\ |\ \|f\|$$

$$\|f + g\| \leq \|f\| + \|g\|,$$

for $f, g \in \mathfrak{F}$ and $\alpha \in \mathbb{R}$. The space $\mathfrak{F}$ endowed with a norm $\|\cdot\|$ is called a *normed linear space.*

Definition 2.2.1. A sequence $\{f_n\}$ in a normed linear space is a *Cauchy sequence* if, for any $\varepsilon > 0$, there exists an $N \geq 1$ such that for any $n, m \geq N$,

$$\|f_n - f_m\| < \varepsilon.$$

Every convergent sequence is a Cauchy sequence.

Definition 2.2.2. A normed linear space $\mathfrak{F}$ is *complete* if every Cauchy sequence converges, i.e., if for each Cauchy sequence $\{f_n\}$ there exists $f \in \mathfrak{F}$ such that $f_n \to f$. A complete normed space is called a *Banach space.*

Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space.

Definition 2.2.3. Let $1 \leq p < \infty$. The family of real valued measurable functions (or rather a.e.-equivalence classes of them) $f : X \to \mathbb{R}$ satisfying

$$\int_X |f(x)|^p d\mu < \infty \tag{2.2.1}$$

is called the $\mathfrak{L}^p(X, \mathfrak{B}, \mu)$ *space* and is denoted by $\mathfrak{L}^p(\mu)$ when the underlying space is clearly known, and by $\mathfrak{L}^p$ where both the space and the measure are known.

The integral in (2.2.1) is assigned a special notation

$$\|f\|_p = \left(\int_X |f(x)|^p d\mu \right)^{\frac{1}{p}},$$

and is called the $\mathfrak{L}^p$ *norm* of f. $\mathfrak{L}^p$ with the norm $\|\cdot\|_p$ is a complete normed space, i.e., a Banach space.

The space of almost everywhere bounded measurable functions on $(X, \mathfrak{B}, \mu)$ is denoted by $\mathfrak{L}^\infty$. Functions that differ only on a set of μ-measure 0 are considered to represent the same element of $\mathfrak{L}^\infty$. The $\mathfrak{L}^\infty$ norm is given by

$$\|f\|_\infty = \text{ess sup}|f(x)| = \inf\{M : \mu\{x : f(x) > M\} = 0\}.$$

The space $\mathfrak{L}^\infty$ with the norm $\|\cdot\|_\infty$ is a Banach space.

Definition 2.2.4. The space of bounded linear functionals on a normed space $\mathfrak{F}$ is called the *adjoint space* to $\mathfrak{F}$ and is denoted by $\mathfrak{F}^*$. The *weak convergence* in $\mathfrak{F}$ is defined as follows: A sequence $\{f_n\}_1^\infty \subset \mathfrak{F}$ converges weakly to an $f \in \mathfrak{F}$ if and only if for any $F \in \mathfrak{F}^*$, $F(f_n) \to F(f)$ as $n \to +\infty$. Similarly, a sequence of functionals $\{F_n\}_1^\infty \subset \mathfrak{F}^*$ converges in the *weak-$*$ topology* to a functional $F \in \mathfrak{F}^*$ if and only if for any $f \in \mathfrak{F}$, $F_n(f) \to F(f)$ as $n \to +\infty$.

Theorem 2.2.1. *Let $1 \le p \le \infty$ and let q satisfy*

$$\frac{1}{p} + \frac{1}{q} = 1\,,\ (\frac{1}{\infty} = 0).$$

Then $\mathfrak{L}^q$ is the adjoint space of $\mathfrak{L}^p$.

If $f \in \mathfrak{L}^p$, $g \in \mathfrak{L}^q$, then fg is integrable and the Hölder inequality holds:

$$\int_X |fg|\, d\mu \le \|f\|_p \|g\|_q.$$

Let $g \in \mathfrak{L}^q$. We define a functional F_g on $\mathfrak{L}^p$ by setting

$$F_g(f) = \int_X fg d\mu$$

$$\|F_g\| = \sup_{f \ne 0} \{\frac{|F_g(f)|}{\|f\|}\}.$$

Clearly F_g is linear.

Proposition 2.2.1. *Each function* $g \in \mathfrak{L}^q$ *defines a bounded linear functional* F_g *on* $\mathfrak{L}^P$ *with* $F_g(f) = \int_X fgd\mu$ *and* $\|F_g\| = \|g\|_q$.

Theorem 2.2.2. *(Riesz Representation Theorem)* [Dunford and Schwartz, 1964, Ch. IV, 8.5]

Let F *be a bounded linear functional on* $\mathfrak{L}^p, 1 \le p < \infty$. *Then there exists a function* g *in* $\mathfrak{L}^q$ *such that*

$$F(f) = \int_X fgd\mu.$$

Furthermore, $\|F\| = \|g\|_q$.

We will use the following kinds of convergence in $\mathfrak{L}^p$ spaces.

(1) Norm (or strong) convergence:

$$f_n \to f \text{ in } \mathfrak{L}^p - \text{norm} \iff \|f_n - f\|_p \to 0,\, n \to +\infty.$$

(2) Weak convergence: $f_n \to f$ weakly in $\mathfrak{L}^p$, $1 \le p < +\infty$, $\iff$

$$\forall g \in \mathfrak{L}^q, \quad \int f_n g d\mu \to \int fgd\mu, \quad \text{where } \frac{1}{p} + \frac{1}{q} = 1.$$

(3) Pointwise convergence:

$$f_n \to f \quad \text{almost everywhere (a.e.)} \iff f_n(x) \to f(x),$$

for almost every $x \in X$.

The following results give several characterizations of these types of convergence and connections between them:

Theorem 2.2.3. *Let a sequence* $\{f_n\}_{n=1}^{\infty}$, $f_n \in \mathfrak{L}^1$, $n = 1, 2, ...$ *satisfy*

(1) $\|f_n\|_1 \le M$ *for some* M;

(2) $\forall \varepsilon > 0\ \exists\ \delta > 0$ *such that for any* $A \in \mathfrak{B}$, *if* $\mu(A) < \delta$ *then for all* n,

$$|\int_A f_n d\mu| < \varepsilon.$$

Then $\{f_n\}$ *contains a weakly convergent subsequence, i.e.,* $\{f_n\}$ *is weakly compact.*

Corollary 2.2.1. *If there exists* $g \in \mathfrak{L}^1$ *such that* $f_n \le g$ *for* $n = 1, 2, ...,$ *then* $\{f_n\}$ *is weakly compact.*

Theorem 2.2.4. *(Scheffé's Theorem)* [Billingsley, 1985] *If* $f_n \geq 0$, $\int f_n d\mu = 1$, $n = 1, 2, \ldots$ *and* $f_n \to f$ *a.e. with* $\int f d\mu = 1$, *then* $f_n \to f$ *in* $\mathfrak{L}^1$*-norm.*

Theorem 2.2.5. *If* $f_n \to f$ *weakly in* $\mathfrak{L}^1$ *and almost everywhere, then* $f_n \to f$ *in* $\mathfrak{L}^1$*-norm.*

We now consider spaces of continuous and differentiable functions. Let X be a compact metric space.

Definition 2.2.5. $C^0(X) = C(X)$ is the space of all continuous real functions $f : X \to \mathbb{R}$, with the norm

$$\|f\|_{C^0} = \sup_{x \in X} \mid f(x) \mid .$$

Definition 2.2.6. Let $r \geq 1$. $C^r(X)$ denotes the space of all r-times continuously differentiable real functions $f : X \to \mathbb{R}$ with the norm

$$\|f\|_{C^r} = \max_{0 \leq k \leq r} \sup_{x \in X} \mid f^{(k)}(x) \mid ,$$

where $f^{(k)}(x)$ is the k-th derivative of $f(x)$ and $f^{(0)}(x) = f(x)$.

Definition 2.2.7. $\mathfrak{M}(X)$ denotes the spaces of all measures μ on $\mathfrak{B}(X)$. The norm, called the total variation norm on $\mathfrak{M}(X)$, is defined by

$$\|\mu\| = \sup_{A_1 \cup \ldots \cup A_N = X} \{\mid \mu(A_1) \mid + \ldots + \mid \mu(A_N) \mid\},$$

where the supremum is taken over all finite partitions of X.

A more frequently used topology on $\mathfrak{M}(X)$ is the *weak topology of measures*, which we can define with the aid of the following result [Dunford and Schwartz, 1964, Ch. IV, 6.3]:

Theorem 2.2.6. *Let* X *be a compact metric space. Then the adjoint space of* $C(X)$, $C^*(X)$, *is equal to* $\mathfrak{M}(X)$.

Definition 2.2.8. The weak topology of measures is a topology of weak convergence on $\mathfrak{M}(X)$, i.e.,

$$\mu_n \to \mu \Leftrightarrow \int_X g d\mu_n \to \int_X g d\mu, \quad \text{for any } g \in C(X).$$

In view of Theorem 2.2.6 this is sometimes referred to as the topology of weak-$*$ convergence.

Theorem 2.2.7. *The weak topology of measures is metrizable and any bounded (in norm) subset of $\mathfrak{M}(X)$ is compact in the weak topology of measures.*

We now present two important corollaries of Theorem 2.2.6.

Corollary 2.2.2. *Two measures μ_1 and μ_2 are identical if and only if*

$$\int_X g d\mu_1 = \int_X g d\mu_2$$

for all $g \in C(X)$.

Corollary 2.2.3. *The set of probability measures is compact in the weak topology of measures.*

For excellent accounts on the weak topology of measures, the reader is referred to [Billingsley, 1968] and [Parthasarathy, 1967].

We now collect a number of results which will be needed in the sequel.

Theorem 2.2.8. [Dunford and Schwartz, 1964, Ch II, 3.6] *Let $\mathfrak{F}, \mathfrak{G}$ be Banach spaces and let $\{T_n\}$ be a sequence of bounded linear operators on $\mathfrak{F}$ into $\mathfrak{G}$. Then the limit $Tf = \lim_{n\to+\infty} T_n f$ exists for every f in $\mathfrak{F}$ if and only if*

(i) the limit Tf exists for every f in a set dense in $\mathfrak{F}$

and

(ii) for each f in $\mathfrak{F}$, $\sup_n |T_n f| < \infty$.

When the limit Tf exists for each f in $\mathfrak{F}$, the operator T is bounded and

$$\|T\| \leq \liminf_{n\to+\infty} \|T_n\| \leq \sup_n \|T_n\| < +\infty.$$

Theorem 2.2.9. *(Rota's Theorem)* [Schaefer, 1974] *If P is a positive operator on $\mathfrak{L}^1(X, \mathfrak{B}, \mu)$, then the set*

$$\{\frac{\lambda}{|\lambda|} : \lambda \text{ is an eigenvalue of } P, \quad |\lambda| = \|P\|\}$$

forms a multiplicative subgroup of the unit circle.

Definition 2.2.8. Let ν and μ be two measures on the same measure space $(X, \mathfrak{B})$. We say that ν is *absolutely continuous* with respect to μ if for any $A \in \mathfrak{B}$, such that $\mu(A) = 0$, it follows that $\nu(A) = 0$. We write $\nu << \mu$.

A useful condition for testing absolute continuity is given by

Theorem 2.2.10. *$\nu << \mu$ if and only if given $\varepsilon > 0$ there exists $\delta > 0$ such that $\mu(A) < \delta$ implies $\nu(A) < \varepsilon$.*

The proof of this theorem can be found in [Dunford and Schwartz, 1964].

If $\nu << \mu$, then it is possible to represent ν in terms of μ. This is the essence of the Radon–Nikodym Theorem.

Theorem 2.2.11. *(Radon-Nikodym) Let $(X, \mathfrak{B})$ be a space and let ν and μ be two normalized measures on $(X, \mathfrak{B})$. If $\nu << \mu$, then there exists a unique $f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu)$ such that for every $A \in \mathfrak{B}$,*

$$\nu(A) = \int_A f d\mu.$$

f is called the Radon–Nikodym derivative and is denoted by $\frac{d\nu}{d\mu}$.

Definition 2.2.9. Let ν and μ be two measures on the same measure space $(X, \mathfrak{B})$. We say that ν and μ are mutually singular if and only if there exist disjoint sets $A_\mu, A_\nu \in \mathfrak{B}$ such that $X = A_\mu \cup A_\nu$ and $\mu(A_\nu) = 0 = \nu(A_\mu)$. We write $\nu \perp \mu$.

Theorem 2.2.12. *(Lebesgue Decomposition Theorem) Let ν and μ be two measures on the same measure space $(X, \mathfrak{B})$. Then there exists a unique decomposition of measure ν into two measures $\nu = \nu_1 + \nu_2$ such that $\nu_1 << \mu$ and $\nu_2 \perp \mu$.*

Definition 2.2.10. Let X be a compact metric space and let μ be a measure on $(X, \mathfrak{B})$, where $\mathfrak{B}$ is the Borel σ-algebra of subsets of X. We define the support of μ as the smallest closed set of full μ measure, i.e.,

$$\text{supp}(\mu) = X \setminus \bigcup_{\substack{\mathcal{O}-\text{open} \\ \mu(\mathcal{O})=0}} \mathcal{O}.$$

It is worth noting that two mutually singular measures may have the same support.

Let $\mathfrak{M}(X)$ denote the space of measures on $(X, \mathfrak{B})$. Let $\tau : X \to X$ be a measurable transformation (i.e., $\tau^{-1}(A) \in \mathfrak{B}$ for $A \in \mathfrak{B}$). τ induces a transformation τ_* on $\mathfrak{M}(X)$ by means of the definition: $(\tau_*\mu)(A) = \mu(\tau^{-1}A)$. Since τ is measurable, it is easy to see that $\tau_*\mu \in \mathfrak{M}(X)$. Hence, τ_* is well defined.

Definition 2.2.11. Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space. Then $\tau : X \to X$ is said to be *nonsingular* if and only if $\tau_* \mu << \mu$, i.e., if for any $A \in \mathfrak{B}$ such that $\mu(A) = 0$, we have $\tau_* \mu(A) = \mu(\tau^{-1} A) = 0$.

Proposition 2.2.2. *Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space, and let $\tau : X \to X$ be nonsingular. Then, if $\nu << \mu$, $\tau_* \nu << \tau_* \mu << \mu$.*

Proof. Since $\nu << \mu$, $\mu(A) = 0 \Rightarrow \nu(A) = 0$. Since τ is non-singular, $\mu(A) = 0 \Rightarrow \mu(\tau^{-1} A) = 0 \Rightarrow \nu(\tau^{-1} A) = 0$. Thus, $\tau_* \nu << \tau_* \mu$. Since τ is nonsingular, $\tau_* \mu << \mu$. □

Definition 2.2.12. Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space. Let

$$\mathfrak{D} = \mathfrak{D}(X, \mathfrak{B}, \mu) = \{ f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu) : \; f \; \geq 0 \text{ and } \|f\|_1 = 1 \}$$

denote the space of probability density functions. A function $f \in \mathfrak{D}$ is called a *density function* or simply a *density*.

If $f \in \mathfrak{D}$, then

$$\mu_f(A) = \int_A f d\mu << \mu$$

is a measure and f is called the *density of* μ_f and is written as $d\mu_f / d\mu$.

Let $\nu << \mu$. We saw in Proposition 2.2.2 that $\tau_* \nu$ is absolutely continuous with respect to μ. Hence the density of ν is transformed into a density of $\tau_* \nu$. This transformation, denoted by P_τ, will be studied in detail in Chapter 4. Clearly $P_\tau : \mathfrak{D} \to \mathfrak{D}$. The operators $\tau_* : \mathfrak{M}(X) \to \mathfrak{M}(X)$ and $P_\tau : \mathfrak{D} \to \mathfrak{D}$ are closely related. Since P_τ acts on $\mathfrak{L}^1$ it is often easier to work with it than with τ_*. The main mathematical tool of this book is P_τ, which is called the *Frobenius–Perron operator* associated with τ.

In Chapter 4, we shall encounter integrals whose analysis is greatly facilitated by a change of variable. Consider the integral

$$\int_c^a f(g(x)) g'(x) dx,$$

where f and g are real-valued functions. We let $u = g(x)$. Then $du = g'(x) dx$ and

$$\int_a^b f(g(x)) g'(x) dx = \int_{g(a)}^{g(b)} f(u) du.$$

We now collect a number of results from functional analysis. Let K be a convex subset of a vector space $\mathfrak{F}$, i.e., for any $g_1, g_2 \in K$, the whole interval $\{tg_1 + (1-t)g_2 : 0 \le t \le 1\}$ is in K. A point in K is called an *extreme point* if it is not an interior point of any line segment lying in K. Thus f is extreme if and only if whenever $f = tg_1 + (1-t)g_2$ with $0 < t < 1$, we have $g_1 \notin K$ or $g_2 \notin K$, i.e., we cannot represent an extreme point as a convex combination of two points in K.

The intersection of all (closed) convex sets containing a set K is the smalest (closed) convex set containing K. This set is called the (closed) *convex hull* of K and denoted by $\mathrm{co}(K)$ ($\overline{\mathrm{co}}(K)$).

Theorem 2.2.13. *(Mazur Theorem) [Dunford and Schwartz, 1964] Let $\mathfrak{F}$ be a Banach space with $A \subset \mathfrak{F}$ where $\overline{A}$, the closure of A, is compact. Then $\overline{\mathrm{co}}(A)$ is compact.*

Theorem 2.2.14. *(Kakutani–Yosida Theorem) [Dunford and Schwartz, 1964] Let $\mathfrak{F}$ be a Banach space and let $T : \mathfrak{F} \to \mathfrak{F}$ be a bounded linear operator. Assume there exists $c > 0$ such that $\|T^n\| \le c$, $n = 1, 2, ...$ Furthermore, if for any $f \in A \subset \mathfrak{F}$, the sequence $\{f_n\}$, where*

$$f_n = \frac{1}{n}\sum_{k=1}^{n} T^k f \ ,$$

contains a subsequence $\{f_{n_k}\}$ which converges weakly in $\mathfrak{F}$, then for any $f \in \overline{A}$,

$$\frac{1}{n}\sum_{k=1}^{n} T^k f \to f^* \in \mathfrak{F}$$

(norm convergence) and $T(f^*) = f^*$.

Theorem 2.2.15. *(Minkowski Theorem) Let K be a closed bounded and convex subset of $\mathbb{R}^n$. Then every boundary point of K is a convex combination of at most n extreme points of K and every interior point is a convex combination of at most $n+1$ extreme points of K.*

2.3 Functions of Bounded Variation in One Dimension

Let $[a,b] \subset \mathbb{R}$ be a bounded interval and let λ denote Lebesgue measure on $[a,b]$. For any sequence of points $a = x_0 < x_1 < ... < x_{n-1} < x_n = b\,, n \ge 1$, we define a partition $\mathcal{P} = \{I_i = [x_{i-1}, x_i) : i = 1, ..., n\}$ of $[a,b]$. The points $\{x_0, x_1, ..., x_n\}$ are called *end-points of the partition* $\mathcal{P}$. Sometimes we will write $\mathcal{P} = \mathcal{P}\{x_0, x_1, ..., x_n\}$.

Definition 2.3.1. Let $f : [a,b] \to \mathbb{R}$ and let $\mathcal{P} = \mathcal{P}\{x_0, x_1, ..., x_n\}$ be a partition of $[a,b]$. If there exists a positive number M such that

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq M$$

for all partitions $\mathcal{P}$, then f is said to be of *bounded variation* on $[a,b]$.

If f is increasing or if it satisfies the Lipschitz condition

$$|f(x) - f(y)| \; < \; K|x-y|,$$

then it is of bounded variation.

Note that the Hölder condition

$$|f(x) - f(y)| < H|x-y|^{\varepsilon}, \; 0 < \varepsilon < 1 \; ,$$

is not enough to guarantee that f is of bounded variation. This can be seen by considering the function

$$f(x) = \begin{cases} x \sin(\frac{1}{x}), & 0 < x \leq 2\pi, \\ 0, & x = 0, \end{cases}$$

which is Hölder continuous, but not of bounded variation (see Problem 2.3.2).

Definition 2.3.2. Let $f : [a,b] \to \mathbb{R}$ be a function of bounded variation. The number

$$V_{[a,b]}f = \sup_{\mathcal{P}}\{\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|\}$$

is called the *total variation* or, simply, the *variation* of f on $[a,b]$.

Many of the following results are well known and can be found in the excellent book [Natanson, 1955].

Theorem 2.3.1. *If f is of bounded variation on $[a,b]$, then f is bounded on $[a,b]$. In fact,*

$$|f(x)| \leq |f(a)| + V_{[a,b]}f$$

for all $x \in [a,b]$.

Lemma 2.3.1. *Let f be a function of bounded variation such that $\|f\|_1 < \infty$. Then $|f(x)| \le V_{[a,b]}f + \frac{\|f\|_1}{b-a}$ for all $x \in [a,b]$, where $\|\cdot\|_1$ is the $\mathfrak{L}^1$ norm on $[a,b]$.*

Proof. We claim there exists $y \in [a,b]$ such that $|f(y)| \le \frac{\|f\|_1}{b-a}$. If not, then for any $x \in [a,b]$

$$(b-a)|f(x)| > \|f\|_1.$$

Hence, $\|f\|_1 = \int_a^b |f(x)| d\lambda(x) > \int_a^b \frac{\|f\|_1}{b-a} d\lambda(x) = \|f\|_1$ and we have a contradiction.

Since

$$| f(x) | \le | f(x) - f(y) | + | f(y) |$$

we have

$$|f(x)| \le V_{[a,b]}f + \frac{\|f\|_1}{b-a}.$$

□

Theorem 2.3.2. *Let f and g be of bounded variation on $[a,b]$. Then so are their sum, difference and product. Also, we have*

$$V_{[a,b]}(f \pm g) \le V_{[a,b]}f + V_{[a,b]}g$$

and

$$V_{[a,b]}(f \cdot g) \le AV_{[a,b]}f + BV_{[a,b]}g,$$

where $A = sup\{|g(x)| : x \in [a,b]\}$, $B = sup\{|f(x)| : x \in [a,b]\}$.

Quotients are not included in Theorem 2.3.2 because the reciprocal of a function of bounded variation need not be of bounded variation. For example, if $f(x) \to 0$ as $x \to x_0$, then $1/f$ will not be bounded on any interval containing x_0 and therefore $1/f$ cannot be of bounded variation on such an interval. To extend Theorem 2.3.2 to quotients, we must exclude functions whose values can be arbitrarily close to zero.

Theorem 2.3.3. *Let $f : [a,b] \to \mathbb{R}$ be of bounded variation and assume f is bounded away from 0; i.e., there exists a positive number $\alpha > 0$ such that $|f(x)| \ge \alpha$ for all $x \in [a,b]$. Then $g = 1/f$ is of bounded variation on $[a,b]$ and*

$$V_{[a,b]}g \le \frac{1}{\alpha^2} V_{[a,b]}f.$$

Proof. Let $\{x_0, ..., x_n\}$ be a partition of $[a, b]$. Since $f \in BV[a, b]$, we have

$$\sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| < M_1$$

Then,

$$\begin{aligned}\sum_{k=1}^{n} \left|\frac{1}{f(x_k)} - \frac{1}{f(x_{k-1})}\right| &= \sum_{k=1}^{n} \frac{|f(x_k) - f(x_{k-1})|}{|f(x_k)||f(x_{k-1})|} \\ &\leq \frac{1}{\alpha^2} M_1.\end{aligned}$$

Therefore, $\frac{1}{f} \in BV[a, b]$ and $V_{[a,b]} \frac{1}{f} \leq \frac{1}{\alpha^2} V_{[a,b]} f$. □

If we keep f fixed and study the total variation as a function of the interval $[a, b]$, we have the following property:

Theorem 2.3.4. *Let $f : [a, b] \to \mathbb{R}$ be of bounded variation and assume $c \in (a, b)$. Then f is of bounded variation on $[a, c]$ and on $[c, b]$ and we have*

$$V_{[a,b]} f = V_{[a,c]} f + V_{[c,b]} f.$$

The following result characterizes functions of bounded variation.

Theorem 2.3.5. *Let f be defined on $[a, b]$. Then f is of bounded variation if and only if f can be expressed as the difference of two increasing functions.*

Theorem 2.3.6. *Let f be of bounded variation on $[a, b]$. If $x \in [a, b]$, let $V(x) = V_{[a,x]} f$ and let $V(a) = 0$. Then every point of continuity of f is also a point of continuity of V. The converse is also true.*

Combining the two foregoing theorems, we have

Theorem 2.3.7. *Let $f : [a, b] \to \mathbb{R}$ be continuous. Then f is of bounded variation on $[a, b]$ if and only if f can be represented as the difference of two increasing continuous functions.*

We now distinguish an important subspace of functions of bounded variation.

Definition 2.3.3. Let $f : [a, b] \to \mathbb{R}$. f is called an *absolutely continuous function* if and only if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any $\{s_i, t_i\}_{i=1}^N$

$$\sum_{i=1}^{N} |t_i - s_i| < \delta \implies \sum_{i=1}^{N} |f(t_i) - f(s_i)| < \varepsilon.$$

If f has a continuous derivative (or more generally, if f is absolutely continuous), there is a very useful representation for its variation.

Theorem 2.3.8. *Let $f : [a,b] \to \mathbb{R}$ have a continuous derivative f' on $[a,b]$. Then*

$$V_{[a,b]} f = \int_a^b |f'(x)| d\lambda(x).$$

We now briefly discuss the interesting relation between absolute continuity (nonsingularity) of a function and nonsingularity of a transformation defined by this function (Definition 2.2.11).

Recall, that $f : [0,1] \to [0,1]$ is called nonsingular (as a transformation) $\Leftrightarrow$ for any $A \in \mathfrak{B}([0,1])$ $\lambda(A) = 0 \Rightarrow \lambda(f^{-1}(A)) = 0$ (i.e. $\Leftrightarrow f_*\lambda << \lambda$ for $f_*\lambda(A) = \lambda(f^{-1}(A))$). Then, by the Radon–Nikodym Theorem, there exists a function $g(x) \geq 0$ such that

$$\lambda(f^{-1}(A)) = \int_A g(t) d\lambda(t), \tag{2.3.1}$$

for all $A \in \mathfrak{B}([0,1])$. Note that the function g may vanish on some set of positive measure. If $f(0) = 0$, then applying the formula (2.3.1) to $A = [0,x]$, we obtain

$$f^{-1}(x) = \int_0^x g(t) d\lambda(t) \quad \text{for} \quad x \in [0,1].$$

On the other hand, the function $\varphi : [0,1] \to [0,1]$ is called nonsingular or absolutely continuous (as a function) $\Leftrightarrow \varphi$ is differentiable a.e., and

$$\varphi(x) = \int_0^x \varphi'(t) d\lambda(t) \quad \text{for} \quad x \in [0,1].$$

(This characterization is equivalent to Definition 2.3.3).

The following proposition is a direct consequence of the definitions:

Proposition 2.3.1. *Let $f : [0,1] \to [0,1]$. If f^{-1} exists and is absolutely continuous as a function, then f is nonsingular as a transformation.*

The following result was proved in [Quas, 1996].

Proposition 2.3.2. *Let $f : [0,1] \to [0,1]$ be a homeomorphism. If f is absolutely continuous as a function and $f' > 0$, a.e., then f is nonsingular as a transformation.*

Proof. For each $x \in [0,1]$, we have

$$f(x) = \int_0^x f'(t) d\lambda(t). \tag{2.3.2}$$

Let $\mu = f' \cdot \lambda$. The measure μ is equivalent to λ. By (2.3.2), we have

$$\lambda(f([0,x])) = \int_0^x f' d\lambda = \mu([0,x]), \quad x \in [0,1].$$

Since the intervals $\{[0,x];\ x \in [0,1]\}$ generate $\mathfrak{B}([0,1])$ we have

$$\lambda(f(A)) = \mu(A)$$

for any measurable $A \subset [0,1]$. Thus, $\lambda(f(A)) = 0 \Leftrightarrow \mu(A) = 0 \Leftrightarrow \lambda(A) = 0$. Since f is a homeomorphism we have $\lambda(B) = 0 \Leftrightarrow \lambda(f^{-1}(B)) = 0$, for any measurable B. This implies that f is nonsingular as a transformation. □

Below we present an example of $f : [0,1] \to [0,1]$ that is nonsingular as a transformation but not absolutely continuous as a function.

Example 2.3.1. Let $\mathfrak{c}$ be the Cantor function (sometimes called the "devil's staircase" [Devaney, 1989]). It is a continuous, increasing function transforming the Cantor set onto $[0,1]$. The derivative of $\mathfrak{c}$, $\mathfrak{c}'$, is equal to 0 almost everywhere. Let $f(x) = \frac{1}{2}(\mathfrak{c}(x) + x)$. Then $f : [0,1] \to [0,1]$ is a homeomorphism. For any measurable $A \subset [0,1]$, we have

$$\lambda(A) \geq \int_{f^{-1}(A)} f'(x) d\lambda(x) = \frac{1}{2}\lambda(f^{-1}(A)).$$

Thus, $\lambda(A) = 0$ implies $\lambda(f^{-1}(A)) = 0$ and f is nonsingular as a transformation. We also have

$$f(1) = 1 > \frac{1}{2} = \int_I f'(x) d\lambda(x),$$

so f is not absolutely continuous as a function.

We now show that the assumption $f' > 0$ a.e. is important in Proposition 2.3.2.

Example 2.3.2. Let f be the homeomorphism of the previous example. f transforms the Cantor set $\mathfrak{c}$ into the Cantor set $\mathfrak{c}_{\frac{1}{2}}$ of measure 1/2. The inverse homeomorphism f^{-1} is absolutely continuous as a function. Its derivative $(f^{-1})'$ is equal to 0 on $\mathfrak{c}_{\frac{1}{2}}$ and equal to 2 elsewhere. It is not difficult to check that, for any $x \in [0, 1]$,

$$f^{-1}(x) = \int_0^x (f^{-1})'(t) d\lambda(t).$$

Obviously, f^{-1} is not nonsingular as a transformation, since $\lambda(\mathfrak{c}) = 0$ and $\lambda(\mathfrak{c}_{\frac{1}{2}}) = \lambda((f^{-1})^{-1}(\mathfrak{c})) = \frac{1}{2}$.

We now present a result due to E. Helly that has many important applications.

Theorem 2.3.9. *(Helly's First Theorem)* [Natanson, 1955]
Let an infinite family of functions $F = \{f\}$ be defined on an interval $[a, b]$. If all functions of the family and the total variation of all functions of the family are bounded by a single number, i.e.,

$$|f(x)| \leq K, \qquad V_{[a,b]} f \leq K \qquad \forall f \in F,$$

then there exists a sequence $\{f_n\} \subset F$ that converges at every point of $[a, b]$ to some function f^* of bounded variation, and $V_{[a,b]} f^* \leq K$.

Two inequalities that will play an important role in the sequel follow:

Theorem 2.3.10. *Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. Let $x, y \in [a, b]$ and $x < y$. Then*

$$|f(x)| + |f(y)| \leq V_{[x,y]} f + \frac{2}{|y - x|} \int_x^y |f(t)| dt.$$

Proof. We have

$$|f(x)| + |f(y)| \leq 2 \inf_{x \leq t \leq y} |f(t)| + |f(x) - f(t)| + |f(t) - f(y)|.$$

By the Mean Value Theorem for integrals, we obtain

$$| f(x) | + | f(y) | \leq \frac{2}{| y - x |} \int_x^y | f(t) | d\lambda(t) + V_{[x,y]} f.$$

$\square$

Theorem 2.3.11. *(Yorke's Inequality)* [Lasota and Mackey, 1985, p. 118]

Let $f : [a, b] \to \mathbb{R}$ be of bounded variation. Let $[c, d] \subset [a, b]$ and let $\chi_{[c,d]}$ be the characteristic function of the interval $[c, d]$. Then

$$V_{[a,b]}(f\chi_{[c,d]}) \leq 2V_{[c,d]}f + \frac{2}{d-c}\int_c^d |f(t)|d\lambda(t).$$

We now make the space of functions of bounded variation into a Banach space. Let

$$BV([a, b]) = \{f \in \mathfrak{L}^1 : \inf_{f_1 = f\, a.e.} V_{[a,b]}f_1 < +\infty\}.$$

Note that the infimum is taken over all functions a.e. equal to f. For example, the function

$$f(x) = \begin{cases} n, & \text{if } x = \frac{1}{n}, \qquad n = 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$$

clearly has infinite variation, but $f \in BV([0, 1])$ since $f_1 \equiv 0$ is a.e. equal to f and $V_{[0,1]}f_1 = 0$.

We define a norm on $BV([a, b])$ as follows: For $f \in BV([a, b])$,

$$\|f\|_{BV} = \|f\|_1 + \inf_{f_1 = f\, a.e.} V_{[a,b]}f_1.$$

Without the $\mathfrak{L}^1$-norm, $\|\cdot\|_{BV}$ would not be a norm, since a function that is not 0 could have 0 variation.

We now collect some miscellaneous properties of $BV([a, b])$.

Proposition 2.3.3. *$BV([a, b])$ is dense in $\mathfrak{L}^1([a, b])$.*

Proof. Since $C^1([a, b])$ is dense in $\mathfrak{L}^1([a, b])$ and $BV([a, b])$ contains $C^1([a, b])$, the result is true. □

Proposition 2.3.4. *A bounded set in $BV([a, b])$ is strongly compact in $\mathfrak{L}^1([a, b])$.*

Proof. If $\{f_\alpha\}_{\alpha \in \mathfrak{A}} \subset BV([a, b])$ is bounded, there exists $K_1 < \infty$ such that

$$\|f_\alpha\|_{BV} \leq K_1 \quad \forall \alpha \in \mathfrak{A}.$$

From the definition of $\|\cdot\|_{BV}$ it follows that $\{f_\alpha\}_{\alpha \in \mathfrak{A}}$ is uniformly bounded, i.e., there exists $K_2 < \infty$ such that

$$|f_\alpha(x)| \leq K_2 \quad \forall \alpha \in \mathfrak{A}.$$

Let $K = \max(K_1, K_2)$. By Helly's Theorem, there exists a subsequence $\{f_{\alpha_k}\}$ such that

$$f_{\alpha_k} \to f^*$$

everywhere. Since $\{f_\alpha\}_{\alpha \in \mathfrak{A}}$ is weakly compact (it is uniformly bounded) and $f_{\alpha k} \to f^*$, Theorem 2.2.5 implies that $f_{\alpha_k} \to f^*$ in $\mathfrak{L}^1([a,b])$. Hence $\{f_\alpha\}_{\alpha \in \mathfrak{A}}$ is strongly compact in $\mathfrak{L}^1([a,b])$. □

Proposition 2.3.5. *If* $V_{[a,b]} f_n \le K$ *for all* n *and* $f_n \to f$ *in* $\mathfrak{L}^1 = \mathfrak{L}^1([a,b])$, *then*

$$V_{[a,b]} f \le K.$$

Proof. Since $f_n \to f$ in $\mathfrak{L}^1([a,b])$, we can assume that some subsequence $f_{n_k} \to f$ everywhere after changing the functions on a set of measure 0. Consider the partition $a = x_0 < x_1 < ... < x_n = b$. Then we have

$$\sum_{i=1}^{N} |f_{n_k}(x_i) - f_{n_k}(x_{i-1})| \le K,\ k = 1, 2, ...$$

Taking the limit as $k \to \infty$, we obtain

$$\sum_{i=1}^{N} |f(x_i) - f(x_{i-1})| \le K.$$

Since the partition was arbitrary, we have $V_{[a,b]} f \le K$. □

Proposition 2.3.6. *Let* $f \in BV([a,b])$ *and assume* $\lambda\{x : f(x) \neq 0\} > 0$. *Let* $\mathrm{supp} f = \{x : f(x) \neq 0\}$ *denote the support of* f. *Then the interior of* $\mathrm{supp}\, f \neq \emptyset$.

Proof. Since f is continuous except at a countable number of points, we can choose x_0 such that $|f(x_0)| = h \neq 0$ and f is continuous at x_0. Since $h \neq 0$, there is a neighborhood of $f(x_0)$, U, such that $0 \notin U$. Since f is continuous at $x_0, f^{-1}(U)$, is open, i.e., $|f(x)| \neq 0$ for x in some neighborhood of x_0. Hence the support of f contains a nonempty open set. □

For n-dimensional functions of bounded variation, this property is not necessarily true [Góra and Boyarsky, 1992].

Below we present two results of [Keller, 1982], which we will use in Section 11.2. Let us define the indefinite integral $\int(\Phi)$ of a function $\Phi \in \mathfrak{L}^1$ by

$$\int(\Phi)(z) = \int_{x \le z} \Phi(x) d\lambda(x).$$

Lemma 2.3.2. *Let $f \in BV$ and $\Phi \in \mathfrak{L}^1$. Then,*

$$|\int f \Phi d\lambda| \le V(f) \cdot \| \int(\Phi)\|_\infty + |\int \Phi d\lambda| \cdot \|f\|_\infty \le 2\|f\|_{BV} \| \int(\Phi)\|_\infty.$$

Proof. Let $J_1, \dots, J_M$ be a partition of $I = [0,1]$ into subintervals $J_i = [a_i, a_{i+1}]$ such that $0 = a_0 < a_1 < \cdots < a_M = 1$ and assume that Φ is constant on each J_i. Let $G = \int(\Phi)$. Then

$$\begin{aligned} |\int f \cdot \Phi d\lambda| &= |\sum_{i=1}^{M} \int_{J_i} f \cdot \Phi d\lambda| = |\sum_{i=1}^{M} u_i \int_{J_i} \Phi d\lambda| \\ &= |\sum_{i=1}^{M} u_i [G(a_i) - G(a_{i-1})]| \\ &\le \sum_{i=1}^{M-1} |u_{i+1} - u_i| \cdot \|G\|_\infty + |G(0) \cdot u_1| + |G(1) \cdot u_M| \\ &\le V(f) \cdot \|G\|_\infty + |G(1)| \cdot \|f\|_\infty \le 2\|f\|_{BV} \cdot \|G\|_\infty, \end{aligned}$$

where $u_i \in \overline{\text{co}} f(J_i)$, the closed convex hull of $f(J_i)$. For a general Φ, the required inequality follows by approximation. □

Theorem 2.3.12. *For $f \in \mathfrak{L}^1$,*

$$V(f) = \sup_{\Phi} \left| \int f \Phi d\lambda \right|,$$

where the supremum extends over all $\Phi \in \mathfrak{L}^1$ with $\| \int(\Phi)\|_\infty \le 1$ and $\int \Phi d\lambda = 0$.

Proof. By Lemma 2.3.2, it follows that

$$V(f) \ge \sup |\int f \cdot \Phi d\lambda|.$$

Hence, we need only prove the reverse inequality. Let $S = \sup_{\Phi} |\int f \cdot \Phi d\lambda|$, the supremum being taken as in the statement of the theorem, and assume that $S < \infty$. Let us choose a sequence $\{\mathcal{P}_n\}$ of finite partitions of I into subintervals, $\mathcal{P}_{n+1}$ finer then $\mathcal{P}_n$, which generates the Borel σ-algebra on I. Then, the conditional expectations $E[f|\mathcal{P}_n] \to f$ a.e.

with respect to λ (see Section 2.4). This implies that for each version $\overline{f}$ of f

$$\frac{1}{\lambda(I_n(x))}\int_{I_n(x)} f d\lambda \to \overline{f}, \quad \text{as} \quad n \to \infty,$$

everywhere except for a set $N(\overline{f})$ of zero λ measure, where $I_n(x)$ denotes the element of $\mathcal{P}_n$ containing x. Now sums of the type $\sum\limits_{i=1}^{k} |\overline{f}(a_i) - \overline{f}(a_{i-1})|$, with $a_i \notin N(\overline{f})$, can be approximated by the integrals $\int f \cdot \Phi d\lambda$, with Φ as required (see Problem 2.3.9). Then,

$$\sup_{\substack{a_0<\cdots<a_k \\ a_i \notin N(\overline{f})}} \sum_{i=1}^{k} |\overline{f}(a_i) - \overline{f}(a_{i-1})| \le S < \infty.$$

That is, $\overline{f}|_{I\setminus N(\overline{f})}$ is of bounded variation and can be extended to a function $\overline{\overline{f}}$ on all of I (by using one-sided limits) such that

$$\sup_{a_0<\cdots<a_k} \sum_{i=1}^{k} |\overline{\overline{f}}(a_i) - \overline{\overline{f}}(a_{i-1})| \le S.$$

Since $\overline{\overline{f}}$ is also a version of f (i.e., $\overline{\overline{f}} = f a.e.$), we finally have $V(f) \le S$. $\square$

2.4 Conditional Expectations

Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space. Let $\mathfrak{C} \subset \mathfrak{B}$ be a σ-algebra. For $f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu)$, we define the conditional expectation of f with respect to $\mathfrak{C}$ as follows:

Definition 2.4.1. Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space and let $\mathfrak{C} \subset \mathfrak{B}$ be a σ-algebra. For $f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu)$, we define the conditional expectation of f with respect to $\mathfrak{C}$ as the Radon–Nikodym derivative of the measure $f\mu_{|\mathfrak{C}}$ with respect to the measure $\mu_{|\mathfrak{C}}$ and denote it by $E(f|\mathfrak{C})$:

$$E(f|\mathfrak{C}) = \frac{d(f\mu_{|\mathfrak{C}})}{d(\mu_{|\mathfrak{C}})}.$$

Theorem 2.4.1. *For a function $g \in \mathfrak{L}^1(X, \mathfrak{C}, \mu)$, we have $g = E(f|\mathfrak{C})$ if and only if*

$$\int_A g d\mu = \int_A f d\mu$$

for any $A \in \mathfrak{C}$.

Example 2.4.1. Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space and let $X = \bigcup_{n=1}^{\infty} A_n$, where $A_n \in \mathfrak{B}$ and $A_n \cap A_m = \emptyset$ for $m, n = 1, 2, ...$ and $n \neq m$. The partition $\{A_n\}_{n=1}^{\infty}$ generates a σ-algebra $\mathfrak{A}$. In this case, for any $f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu)$, we have

$$E(f|\mathfrak{A}) = \sum_{n=1}^{\infty} \frac{1}{\mu(A_n)} \int_{A_n} f d\mu \cdot \chi_{A_n}.$$

It is easy to check that the right hand side of the above equality satisfies the condition given in Theorem 2.4.1.

Conditional expectations have all the properties of integrals. Some other basic properties of conditional expectations are listed in the following:

Theorem 2.4.2. *Let $(X, \mathfrak{B}, \mu)$ be a normalized measure space.*
(a) If $\mathfrak{C}_1 \subset \mathfrak{C}_2 \subset \mathfrak{B}$ and $f \in \mathfrak{L}^1(X, \mathfrak{B}, \mu)$, then

$$E(E(f|\mathfrak{C}_2)|\mathfrak{C}_1) = E(f|\mathfrak{C}_1).$$

(b) If $\mathfrak{C}_n \subset \mathfrak{B}$, $n = 1, 2, ...$, is an increasing sequence of σ-algebras ($\mathfrak{C}_n \subset \mathfrak{C}_{n+1}$, for any $n \geq 1$) and $\mathfrak{C} = \sigma(\bigcup_{n \geq 1} \mathfrak{C}_n)$, then

$$E(f|\mathfrak{C}_n) \longrightarrow E(f|\mathfrak{C})$$

μ-a.e. and in $\mathfrak{L}^1(X, \mathfrak{C}, \mu)$.
(c) If $\mathfrak{C}_n \subset \mathfrak{B}$, $n = 1, 2, ...$, is a decreasing sequence of σ-algebras ($\mathfrak{C}_n \supset \mathfrak{C}_{n+1}$, for any $n \geq 1$) and $\mathfrak{C} = \bigcap_{n \geq 1} \mathfrak{C}_n$, then

$$E(f|\mathfrak{C}_n) \longrightarrow E(f|\mathfrak{C})$$

μ-a.e. and in $\mathfrak{L}^1(X, \mathfrak{B}, \mu)$.

Problems for Chapter 2

Problem 2.2.1. Prove Scheffé's Theorem (Theorem 2.2.4).

Problem 2.3.1. Let $f(x) = \sin(x), \quad x \geq 0$.
(a) Prove that $f \in BV[0, 2\pi]$.
(b) Find $V(x) = V_{[0,x]} f$, for any $x > 0$.

Problem 2.3.2. Prove that the function $f : [0, 2\pi] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} x \ \sin\left(\frac{1}{x}\right), & \text{if } 0 < x \leq 2\pi \\ 0, & \text{if } x = 0 \end{cases}$$

is not of bounded variation on $[0, 2\pi]$.

Problem 2.3.3. Let

$$f(x) = \begin{cases} x^2 \ \sin\left(\frac{1}{x}\right), & \text{if } 0 < x \leq 2\pi \\ 0, & \text{if } x = 0. \end{cases}$$

Prove that $f \in BV[0, 2\pi]$.

Problem 2.3.4. Prove that if $f \in BV[a, b]$, then f is bounded on $[a, b]$.

Problem 2.3.5. Suppose $f, g \in BV[a, b]$. Prove that $fg \in BV[a, b]$.

Problem 2.3.6. Let f, g be functions of bounded variation, $g(x) \geq \sigma > 0$. Prove that $\frac{f(x)}{g(x)}$ is of bounded variation.

Problem 2.3.7. Let $f : [a, b] \to \mathbb{R}$ satisfy a Lipschitz condition. Prove that $f \in BV[a, b]$.

Problem 2.3.8. Let f have support in $[b, c]$ and let it be of bounded variation. Let g have support in $[-a, a]$ and $\int_{-a}^{a} |g(t)| d\lambda(t) \leq 1$. Prove that

$$V_{[b-a,c+a]}(f * g) \leq V_{[b-a,c+a]} f \ ,$$

where $*$ denotes convolution:

$$f * g(t) = \int_{-\infty}^{+\infty} f(s) g(t - s) d\lambda(s) = \int_{-\infty}^{+\infty} g(s) f(t - s) d\lambda(s),$$

$t \in \mathbb{R}$.

Problem 2.3.9. Let $f \in BV[0, 1]$ and let points $0 = a_0 < a_1 < \cdots < a_k = 1$ be given. Construct a sequence of functions $\Phi_n \in \mathfrak{L}^1$, $\| \int(\Phi_n)\|_\infty \leq 1$, $\int \Phi_n d\lambda = 0$, $n = 1, 2, \ldots$, such that the integrals $\int f \Phi_n d\lambda$ approximate $\sum_{i=1}^{k} |f(a_i) - f(a_{i-1})|$.